DOCUMENT RESUME

ED 046 064 EA 003 183

AUTHOR Brain, George B.
TITLE National Assessment -- Evaluation and Accountability.
PUB DATE Oct 70
NOTE 14p.; Paper presented at National Association of State Boards of Education meeting. (October 1970)

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Comparative Testing, Educational Objectives, Educational Planning, Educational Testing, *Evaluation, *Evaluation Methods, *Evaluation Techniques, Measurement, *Measurement Techniques, Performance, Student Behavior, Student Testing, Testing Programs

ABSTRACT

Most standardized testing programs do not furnish examples of the kinds of information young people actually know. Instead they indicate only (1) how far an individual student is above or below an average score, and (2) the average score of a classroom or a school in relation to others. In contrast, National Assessment evaluates educational changes within the larger population, its purpose being to provide more adequate information about educational quality on a regional and a national basis. Based on a model of statistical sampling similar to public opinion polls, this program describes the knowledge, skills, attitudes, and understandings of groups of persons at the ages of 9, 13, 17, and young adults. National Assessment measures "population knowledge," rather than the knowledge of an individual. Results for science and citizenship show two trends: (1) knowledge of students increases with age, and (2) adults tend to forget much of what they learned in school. (Author/LLR)

ED 046064

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

NATIONAL ASSESSMENT -- EVALUATION AND ACCOUNTABILITY

By
George B. Brain

From the puberty rights of the upper Paleolithic period to present-day forms of objective tests, evaluation and assessment have played important roles in educational planning. Man has always felt the need of defining certain characteristics that were needed for successful functioning. Conforming behavior in a social setting became the first major criterion for this kind of assessment. Certain present or past characteristics were deemed to be predictive of future congruence to the criterion.

Though the early predictors were variously assessed in terms of manliness, hunting prowess, knowledge of sexual mores, and recognition of authority they emerged in early socities as meaningful dimensions of evaluation or assessment.

The more specific criteria of effective human conduct were basically related to the stable predictable patterns of behavior which were deemed consonant with the survival of the group. Eventually, certain kinds of behavior developed into powerful means of education. So early man in his own trial and error method found that certain characteristics that were observable or measurable were reliable predictors of stable group behavior.

In order to insure the criteria of effective and predictable human behavior, a number of mechanisms of control emerged. These mechanisms had as their goal the collective security of the group. They are now identifiable as the family, customs, laws, chieftain's religious beliefs, and education. To early man, they represented a concrete framework by which the cultural heritage was transmitted and the survival of the group insured.

EA 003 183

The earliest form of assessment was global in nature. Though no attempt was made to separate intellectual functioning from personality, the behavioral patterns of the individual were subjected to the criterion of group collective security.

The early leaders of men were empirical psychologists. Homer and the Bible provide examples of the psychological wisdom of early men.

Ulysses on one occasion, not wishing to go to war, pretended to be mad. He began to plow a field sowing salt. One of his friends placed his young son directly in front of the plow; and Ulysses turned aside, showing indeed that he was not mad and could not escape fulfilling his promises.

The Bible, too, is filled with descriptions of how behavior was assessed. In the account of Genesis, Adam hid himself and God is portrayed as being aware that something was wrong because of this change in Adam's behavior.

Likewise, in one of the early tribal wars, Gilead found a way to distinguish between friendly and unfriendly groups. He asked them to pronounce the word "Shibboleth." If they could not pronounce it correctly, they were judged the enemy.

Thus the oral examination plus prescribed feats of courage served as a predictor of effective behavior among early men. Later on as formal education developed, the degree of memory recall and conformity to school expectations also became a part of the evaluation procedure.

But it was not long before a theoretical framework was joined to the observation process. Plato in his hierarchy of society demarcated certain kinds of behavior expected by various levels of society. He expected the

philosopher rulers to manifest verbal ability and intellectual wisdom, the soldiers to evidence courage and manliness, and the vast masses to be concerned only with bodily functions.

Hippocrates discerned differences in men's constitutional make-up suggesting that all individuals fell into the categories of phlegmatic, choleric, sanguine, or melancholic.

Within the nineteenth century, further developments took place in the assessment of behavior. An extension of the oral examination was developed in the essay examination. During the present century, methods of measurement used in the physical and biological sciences have been gradually introduced into the study of human behavior. This approach has been sparked by a desire to become more objective in the analysis of human behavior.

Accompanying the expansion of testing programs, a new technological development, the use of scoring machines and computers, has made the analysis of data easy and economical to obtain.

The fantastic developments in computer technology and data processing illustrate dramatically the new dimensions of assessment procedures. It is possible to compare individuals or groups of individuals with large samples and norms developed from these large samples.

But though the procedures have become exceedingly advanced it is quite apparent that evaluation of pupil behavior moved away from the original intent of assessment procedures.

The early procedures were related to making judgments about the predicted success or failure of an individual within a specific cultural setting.

The recent direction of evaluation, despite its technical and methodological excellence, subtly introduced a new criterion: comparison with the average.

Explicity or implicitly, the test score became the criterion in itself rather than a means to an end. The mathematical symbol became all too often the criterion of expected performance without any real reference to the behavioral phenomenon needed for success. The purpose of the National Assessment design is to return to the idea of making judgments about groups of individuals within specific cultural settings.

Man's earliest sociological groups held that certain stable patterns of behavior were necessary to maintain the security of the group, engage in collective food gathering, and provide for mutual protection. The extended family, customs, law, ethics, religion, priesthood, chieftains, police functions, and education all seem to have developed as man's unique response to his environment. These controls, however, were chiefly imposed on man from without. The individual had little choice but to conform to these external control mechanisms. Another bolstering factor in this control was the fact that the family was often anchored to the land and dependent on an agrarian economy.

Today the characteristics of life in America have changed radically. A 1960 survey showed that one out of five Americans changed his place of residence within that year. As a result the typical control mechanisms of the past have weakened, giving rise to the charge by many that there is a sense of cultural drift in American values.

Mobility today means that in most instances the family exists alone and independently in a strange area. Very seldom do uncles, aunts, or cousins,

grandparents, and others reside next door or even in the same geographical region. Mobility means that the significance of property as a psychological anchor and root has also been changed. Banks indicate that the average length of occupancy of a house with a 25-year mortgage is 7 years.

Changes have resulted in sex roles with women coming to marriage not with feather beds, cooking ability, and canning prowess but with professional competencies, such as teaching, nursing, or secretarial work.

Mobility has resulted in changing patterns of housing, such as are seen in the vast suburban developments where relatively homogeneous groupings are imposed by qualifying salary criteria and contracting agreements.

Changes in the strength of institutional religion have also occurred. It is a well-known fact that a large part of our population changes its religious affiliation because of marriage or personal preference.

Methods of child rearing have also changed with accent on more permissive patterns. From more direct controls, the family has learned to rely on more indirect ones. Boys get less punishment and girls less indulgence and protectiveness. Though girls are possibly less debilitated by parental affection and power, boys find it more difficult to identify clear roles to envy or emulate. The end result may be that the procedures themselves are less effective in the development of leadership and responsibility.

To all of these changes must be added the continual crisis orientation of modern political developments with the need for competitive achievement in the areas of space satellites, moon probes, and atomic arsenals for defense.

The international scene for the past 20 years has been one of continued peaks of anxiety about global destruction and warfare.

Federal programs often in response to international demands or conditions have affected the huge segments of our population through educational programs and defense contract spending. Meanwhile riots and demonstrations have marked the difficult struggle of American citizens to deal with the adaptations required. The cultural anchors of man's psychological security have been uprooted and man's challenge to the alternative of being controlled from external sources has produced the necessity of developing controls from within. Essentially, man's social and psychological orientation has not kept pace with the changes that are occurring all too rapidly in the dimensions of technological change.

Thus in my view, assessment once again has emerged as a sort of cultural control mechanism, but there has been added to it the idea of public accountability.

Assessment investigators change over a period of time for the various subject areas included in the assessment program. And that in and of itself can prove to be its most significant contribution.

Since public education is in the political domain, many decisions affecting education are made on the basis of the self-interest of the decision makers. A vote in the General Assembly; a judgment by the Governor; decisions by superintendents, teachers, and principals; votes in school board meetings; and resolutions and actions of educational organizations often tend to reflect interests other than the prime constituency of learning -- students in schools. Where decisions in public education do not reflect some genuine attempts to move

students more substantially toward goals of educational quality, then the publics served by the schools increasingly will demand public accountability for those decisions.

Obviously there are difficulties in measuring relationships between educational decisions and quality in education. However, the trend is toward development of educational goals, assessment of student performance and evaluation of effective and qualitative school services and processes. This sequence is becoming more concrete all the time, and it will be used to demand accountability from those who make decisions in public education.

So now the schools must deal with evaluation, assessment, and accountability. Any accountability program should seek to develop a consensus on educational goals and objectives. It should identify the school services and processes which can advance students toward objectives and goals as well as those which impede student achievement. From this sequence of goals-objectives-services and processes, schools should develop a system which can be used at all levels of educational decision making to appraise those decisions within the framework of goals and means to those goals. The central question should be "to what extent are decisions affecting school services and processes advancing or impeding student growth and changes in behavior with respect to specific objectives and broader goals?"

Increasingly the American public is becoming output oriented. Citizens realize that the success of many things can be measured by specific outcomes--such as the landing of an Apollo flight on the moon--or business success through investment growth or profits. Obviously we in education are a long way from

knowing what we should know about relationships among school services, processes, achievement and goals, not to mention the impact of out-of-school student life and environment on learning. But the time has come to make quality in education a truly operational force for improving the quality of mankind. Evaluation, Assessment or Accountability must not be viewed with alarm or concern by school boards or school officials, but as tools to be employed with professional precision for improving the quality and the output of the public educational enterprise for which they are legally responsible. We are entering an era when every institution, public or private, will be held to a new level of accountability by federal and state agencies of government working at new levels of sophistication.

Most school standardized testing programs in use today do not furnish examples of the kinds of information young people actually know, instead they indicate: (1) how far a particular student is above or below an average score; and (2) how far the average score of classroom or school is above or below others with which it is compared.

National Assessment is designed to describe the knowledge, skills, attitudes and understandings of persons in four selected age groups. The results consequently sample what these young people know and do not know. Assessment employs statistical sampling procedures to determine what certain groups of people know, rather than what certain individuals know. National Assessment is a measure of "population knowledge" not the knowledge of an individual.

A standardized test is a different device. The usual test, standardized or otherwise, is designed to yield a score for each individual who takes the test.

That score indicates where a student ranks in relation to other students, how he rates in relation to a norm or standard.

The customary standardized test is constructed to measure the ability or achievement of an individual, and also can be used to compare classes or schools.

Standardized tests are extremely useful and essential devices for teachers, administrators and other educators. National Assessment is not a substitute for these education functions. On the other hand, National Assessment is an evaluation of educational changes within the larger population, and its purpose is to provide more adequate information on the quality and success of educational efforts on a regional and national basis.

To know that Johnny scores at the 90th percentile on a standardized reading test is perhaps interesting to Johnny and to Johnny's parents, teacher and school; but it tells nothing about the specific kind of material that Johnny can actually read or understand. Standardized tests do not provide a descriptive analysis of the actual kinds of knowledge or skills young people, as a whole, have or do not have.

National Assessment was specifically designed not to provide a score for each individual. On the contrary, assessment was created on the model of statistical sampling such as is used in public-opinion polls, to establish a census of actual knowledge of groups of persons in the United States at the ages of 9, 13, 17, and young adults. Not every individual must participate in the assessment. Because the sample is chosen on statistically sound grounds, it is possible to generalize from the response of the individuals to the larger group which they represent.

National Assessment differs from standardized tests in other ways. National Assessment uses a greater variety of exercise formats. For example, there were various kinds of equipment used in the exercises on Science. The students used these devices to demonstrate their knowledge and skill in using scientific equipment. The assessment in Writing actually involved writing, not just asking questions about writing practices and techniques. The Citizenship assessment included evaluation of students in group discussions.

Another difference is that while standardized tests are generally administered so that a student reads each question to himself, the National Assessment model provides for a taped voice to read all instructions and exercises while the student reads along in the booklet in front of him. This method was used with groups of about 12 students who were assembled to take the assessment package. A few packages were administered to individual students, and an interviewer then read each exercise aloud and wrote down the student's response. For the adult assessment, each adult was interviewed in his home by a trained interviewer.

Finally, the method of administration of assessment is different from the usual test. National Assessment hired its own field staff to contact schools, draw the sample of students or adults, and administer the packages of exercises. No school personnel were involved in the administration of the exercises to students, other than to assist with basic arrangements for the assessment. The method of hiring a permanent field staff, plus using taped reading of instructions and exercises, increased the comparability of the assessment procedure used all over the country, as well as lessening any burden a local school might feel if

asked to conduct the administration of exercises itself. Moreover, it added to the credibility of the process.

There are a number of immediate benefits from the National Assessment program. National Assessment has provided:

1. The first national description of samples of learning behavior among selected populations of the nation. These tell how as a nation we stack up educationally.
2. Clarification of the difference between the traditional testing for sorting purposes and evaluating the attainment levels of selected population groups. National Assessment provides a new model for assessing education attainment.
3. Demonstration that educational objectives of a school can be formulated and agreed upon in a way that involves parents, laymen and school staffers with a resultant increase in common understanding of what objectives the schools are attempting to accomplish. National Assessment has provided one workable model for citizen involvement in setting educational objectives.
4. Demonstration that assessment exercises can be constructed to provide information about the entire range of school children, not merely data about the average child. National Assessment indicates what the lowest and highest ability ranges know.
5. Demonstration of the feasibility and desirability of using a variety of techniques (e.g. questionnaires, interviews, and performance, etc.) rather than just traditional paper and pencil tests. In the past it has

been believed that every pupil must be given every appraisal exercise and hence too much time would be required if several techniques were used, but National Assessment employs a method which divides exercises into booklets, or packages so that no one pupil takes more than a fraction of the total; yet, achievements of a total pupil group can be measured. The techniques of assessment administration (i.e. written material plus tape recording) were designed to minimize reading limitations.

6. In addition to these direct benefits of National Assessment, a number of "by product" benefits can be utilized by local school leaders. For example, since National Assessment provides specific information on the strengths and weaknesses in educational achievement, the program design should contribute to more effective curriculum development and assist local school officials in developing programs of emphasis within their school system.

7. National Assessment will provide a substantial source of data useful for further research into and understanding of the problems of the central cities, suburbs, differing geographic areas, and communities, as well as Black and non-Black, etc. National Assessment is a new and important collection of information about the United States which will have many diverse applications.

Two major trends--one comforting and the other disturbing--emerge from analysis of the answers for science and citizenship, the only two areas reported so far:

1. The knowledge of students increases with age.

2. Adults tend to forget much of what they learned in school.

In the vast majority of exercises in both subject areas, the percentage of respondents answering correctly rose from age 9 thru 13 to 17. And on other questions, mainly technical ones in science, fewer adults than 17-year-olds gave the correct answer.

For instance, more 13-year-olds than 9-year-olds picked 70 degrees as the most comfortable temperature for a school room, and more 17-year-olds than 13-year-olds knew that the Senate is oneof the two houses of Congress.

But in the science area, adults did not do as well as 17-year-olds in such things as identifying Charles Darwin as the author of the theory of evolution by natural selection.

The major exceptions were on questions that involved practical knowledge rather than book learning. Adults outscored students whenever knowledge had to be applied in a practical context.

There were items that most respondents knew, like where babies come from and the name of the President.

Few trends were identifiable but the most striking were the tendency of children with educated parents to do best, and the fact that black students performed as well or nearly as well as whites on the science questions.

Many of the exercises on citizenship were designed to gauge attitudes and opinions rather than knowledge.

Answers to one series of questions showed that adults tend to be apathetic in civic action. For instance, 86 per cent cited ways in which citizens can

influence government, but only 61 per cent thought they could influence their own state officials. And 61 per cent could think of reasons to write public officials, but only 27 per cent said they had done so.

More 17-year-olds than adults felt they should act to stop an instance of racial discrimination, and more 13 and 17-year-olds said they would be willing to have someone of another race as a neighbor or doctor.

National Assessment will, it is hoped, raise a great many questions among both educators and interested citizens. Aside from such a general and vital question as "Where do we stand, where are we going, where should we plan to go?" A direct benefit of the Assessment is the posing of a number of questions for exploration by the education community. Among them:

1. Should state and local school systems make their own assessments?
2. Will the states find it desirable, practical or politic to develop programs for interstate comparisons?
3. Should test publishers make assessment-like instruments available to states and local schools?
4. Can the National Assessment model, using its sampling techniques, experiences in objectives and exercise formulation, be adapted to smaller geographical units such as a school district?
5. Do some subjects, such as reading and mathematics, lend themselves to "national objectives" without posing a threat to local control of curricula?
6. If National Assessment data prove meaningful should Assessment activities (the evaluation of other subject areas) be speeded up (at considerable increase in costs)?